Sci-Tech News

Volume 47 | Issue 3 Article 4

August 1993

Interconnecting and Accessing Information in the 1990s

Everett Brenner

INTERCONNECTING AND ACCESSING INFORMATION IN THE 1990s

By Everett H. Brenner

In the Spring of 1992 I wrote that we had all been witness to the failure of Communism, but that we have also been witness to the adverse reaction to Margaret Thatcher and George Bush; that "laissez-faire" has had its successes but that perhaps one needs a little less "laissez" in these times; that perhaps a little more centralized power may be helpful; that cooperation doesn't necessarily lead to regulation and lack of control; and that perhaps more centralization and friendly cooperation will be the trend for the next ten to twenty years.[1] At that time I was favoring increased government activity and the building of partnerships in the information industry as a whole. My feeling was that investment money for developments needed in our industry was tight all over, and, therefore, significant progress in retrieval systems for the long run could be made only by subsidized projects and profitable cooperative efforts.

In just over a year, there have been some interesting new alliances in the information industry, and the activity in the government reminds one a little of the post-war years of the 1950s and early 1960s. The events taking place in the government sector may hopefully lead to real change leading to real progress. That kind of change would lead to transformation, not merely modification or even innovation. So let's analyze a bit more closely activities at the National Technical Information Service (NTIS), the National Aeronautics and Space Agency (NASA), the Library of Congress (LC), and the Internet (wherever that may be).

NTIS, under a new management, is moving from a passive to a proactive operation, will focus on customer service, will benchmark costs against the private sector, and will just plain manage their operating systems better. NTIS will no doubt become more effective in promulgating government information, but if it becomes too entrepreneurial the Industrial Information Association (IIA) will get on its back as they have done before. NTIS is attempting to **modernize** its systems and perhaps develop some **innovative** marketing, but none of this activity represents **transformation**, that which would have long-term effects on the industry.

NASA is changing its methods and priorities and, as in the case of NTIS, the focus will be on the customer. NASA's approach is to modernize, literally to catchup. It is 20 years behind the times. It is ironical that NASA was responsible for RECON in the 1950s, one of the most

influential developments (DIALOG) in the field in subsequent years, and here it is just trying to catch up. What we are looking for is a subsidized development with as transforming an effect as was the result of RECON. We won't find it at NASA. The Sputnik pressure of the 1960s is gone, and no equivalent pressure exists.

The change in approach at the Library of Congress is hard to understand. Its attempt to become entrepreneurial, with all the problems related to copyright, etc., can hardly satisfy anyone, but perhaps it's just insane enough to appeal to the Congress. I hope not. This does not represent modification, innovation, or transformation, just provocation.

Internet is in a league all its own. This is where we can expect transformation. I never thought I would ever say that a project as messy and chaotic, lacking the kind of quality I prize, may be a means to real progress. The reality is that this supposedly free network has attracted a host of users (perhaps because it is perceived as free), mostly through its bulletin board groups (egghead groupies?). Because of the amount of traffic on the network, pressure will build to create some order out of the chaos. This will translate into means for navigating through large amounts of raw information which should mean establishing innovative retrieval systems and links, so important for hypertext success. Although "intelligent" navigators, known as KNOWBOTS, are amusingly referred to as DUMBOTS in their present state of development on Internet, it is significant that a beginning has been made in applying intelligence to the browsing process of information retrieval.

This leads me to the most significant issue I believe we face in hoping for transformation for information retrieval systems, that is, the change from retrieval software offering Boolean searching only to software offering multiple means of access. For an article, conceived in April and published in June, I had been asked to focus on an area I consider my expertise and predict its development by the year 2000. I predicted there would be little meaningful progress in the area of human/machine interface problems by the year 2000, and that if I lived to that year, I hoped to eat crow. Although I still stick to that prediction, in just one month I have seen a glimmer of hope which could conceivably lead to a change in my eating habits in the year 2000.[2]

In May at the National Online Meeting in New York the keynoter, Carl Frappaola of the Delphi Group emphasized the importance of "information access". Rarely is that a major part of a keynoter's presentation. He spoke of: Query-by example (intuitive searching, heuristic association, statistical techniques); document clustering (document level searching, customization, virtual areas, trends analysis); semantic networks (syntactics, statistics, morphology); character-based indexing and retrieval (interlanguage query); information retrieval (pattern recognition, neural networks).

But a keynote speech is just "talk." We're now getting a bit of "action," as well. For example, the West Publishing Co's Westlaw's WIN is basically a statistical weighted/relevance system software. This kind of Salton-like system has been around a long time but has never been applied to as large a database

as Westlaw offers. This is a sign that competitive pressure has arrived, and, since Westlaw is a large database, this non-Boolean approach may finally get its chance for development and refinement. PERSONAL LIBRARIAN:CONCEPT SEARCHING (PL) from Personal Library Software is another example of this retrieval approach to searching raw text. This search software appears to be more sophisticated than WIN in that after finding a most relevant document to a query one can automatically adjust one's strategy based on the keywords of that document, thus achieving "document-by-example" retrieval. A most important example of the emergence of non-Boolean approaches is ConQuest from Conquest Software. Significant with this software is the use of semantics in a commercial system. It is a text search and retrieval system which uses published dictionaries to index data automatically. The ConQuest people seem to be taking advantage of a few different access approaches.

I like what I see—A large database taking weighted/relevance systems seriously; software which allows better navigation through a document-by-example approach; and software utilizing multiple access approaches. Advances in this area will be instrumental in catching up with technology we are underutilizing. But—a warning! We have to be patient. These approaches are in a developmental stage. Just as Boolean searching has improved over time because of improved software based on usage and experience, so must we allow for development for these non-Boolean approaches. The rewards will be great, but we'll complain a lot in the process.

Thus—interconnecting and navigating through Internet is certainly the significant activity to follow in the 1990s. Morry Goldstein of Information Access said recently he was really looking forward to Internet II. Hopefully, serious development in the area of automatic retrieval will accompany all that will be learned from interconnecting, and we will look back on the 1990s as truly innovative and years of significant change (transformation).

References

[1]Brenner, E. *Information insights: the road to knoware.* Oxford and New Jersey: Learned Information; 1992; p. 280

[2]Brenner, E. Who knows what the year 2000 will bring? *Online and CD-ROM Review.* 17(2)1993;102-103.

Everett Brenner is located at 43 Sandy Hollow Road, Port Washington, NY 11050.